

Polynomial Bell inequalities

Rafael Chaves^{1,2}

¹*Institute for Physics & FDM, University of Freiburg, 79104 Freiburg, Germany*

²*Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany*

It is a recent realization that many of the concepts and tools of causal discovery in machine learning are highly relevant to problems in quantum information, in particular quantum nonlocality. The crucial ingredient in the connection between both fields is the tool of Bayesian networks, a graphical model used to reason about probabilistic causation. Indeed, Bell's theorem concerns a particular kind of a Bayesian network and Bell inequalities are a special case of linear constraints following from such models. It is thus natural to look for generalized Bell scenarios involving more complex Bayesian networks. The problem, however, relies on the fact that such generalized scenarios are characterized by polynomial Bell inequalities and no current method is available to derive them beyond very simple cases. In this work, we make a significant step in that direction, providing a general and practical method for the derivation of polynomial Bell inequalities in a wide class of scenarios, applying it to a few cases of interest. We also show how our construction naturally gives rise to a notion of nonsignalling in generalized networks.

Bell's theorem [1] demonstrates that our classical conceptions of causal relations must be taken with care, as they fail to commit with the results obtained in some quantum experiments performed by distant parties, the phenomenon known as quantum nonlocality. Even without detailed information about the underlying processes, the causal structure of the setup alone already implies strong constraints – the famous Bell's inequalities – on the correlations that are compatible with it.

This is close to the reasoning employed in the field of causal inference [2, 3], a connection that has recently attracted considerable attention [4–12]. Since Bell's theorem is a statement about classical correlations, it comes as no surprise that mathematical tools and concepts, originally devised in a causal inference context, can also be applied to the study of nonlocality. Indeed, Bell's theorem concerns the same kind of causal structures that are the object of study in Bayesian networks [2] and Bell inequalities are a special case of linear constraints following from such models [4]. Bayesian networks not only offer a new conceptual perspective to revisit quantum nonlocality [8, 11] but also provide the right language to devise generalized Bell scenarios [5, 6].

Several extensions of the paradigmatic Bell experiment – two distant parties, performing two possible experiments on their shares of a joint system – have been proposed, including more parties [13, 14], more measurements/outcomes [15, 16] and sequential measurements [17, 18]. However, all these different generalizations share the same basic property: the correlations between all the parties originate from a single (not directly observable) source, being therefore named as local hidden variable (LHV) models. In spite of the rich plethora of phenomena and applications [19], LHV models represent a very particular case of the possibilities offered by Bayesian networks. Those typically include several

independent hidden variables and will be named here as generalized local hidden variable (GLHV) models. These scenarios with many independent sources are also ubiquitous in quantum information, e.g., entanglement percolation [20], entanglement swapping [21] and quantum repeaters [22, 23]. Thus, understanding generalized Bell scenarios is not only of fundamental interest but also of high practical relevance.

Within that context, the basic question to be solved is how to derive Bell inequalities for general Bayesian networks. Bell inequalities play a fundamental role in study of nonlocality, since it is via their violation (e.g. with quantum entangled states) that we can witness the nonlocal character of a given experimental data. Unfortunately, as opposed to usual Bell scenarios, a GLHV model implies a non-convex region – characterized by polynomial Bell inequalities – of correlations that are compatible with it. Generally, algebraic geometry methods can be used to characterize such polynomial constraints [24, 25], but given their computational complexity, in practice they are intractable even for very simple models [4]. Arguably, because of this difficulty, only sparse results have been obtained in the derivation of Bell inequalities for GLHV models, either using coarse-grained information [5, 9, 26–28] or considering particular scenarios [29–32]. However, to our knowledge, no practical and systematic method for the derivation of polynomial Bell inequalities for GLHV models is known to this date.

In this paper we propose a general method for deriving polynomial Bell inequalities in a wide class of Bayesian networks. In spite of the non-convex character of the problem, we show how to obtain polynomial inequalities resorting to a linear programming technique, namely a Fourier-Motzkin elimination [33]. We illustrate the general method applying it to a few relevant cases and derive new polynomial Bell inequalities. Fur-

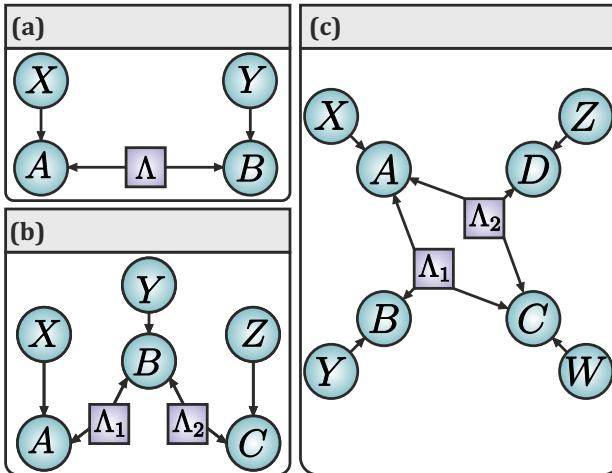


FIG. 1. DAG representation of Bayesian networks. (a) Bipartite LHV model. (b) GLHV model with 2 independent hidden variables representing the bilocality scenario of [29]. (c) GLHV model with 2 hidden variables shared among 4 parties.

thermore, we explain how our construction naturally leads to a notion of nonsignalling correlations [34] in generalized Bell networks.

I. BELL INEQUALITIES, BAYESIAN NETWORKS AND MARGINAL PROBLEMS

Bell scenarios beyond LHV models can be represented via the graphical notation of Bayesian networks [2, 3]. Underlying models are represented by directed acyclic graphs (DAG), where nodes stand for variables and directed arrows represent their causal relations [2]. While LHV models correspond to a DAG with a single hidden variable (see Fig. 1(a)), GLHV models are represented by DAGs with $n \geq 2$ independent hidden variables (see Fig. 1(b)-(c)). The causal relations implied by a DAG are captured by the (conditional) independencies (CI) implied by the graph and that can be listed by the d-separation criterion [2]. For instance, for the LHV model in Fig. 1(a) it follows that $p(x, y, \lambda) = p(x)p(y)p(\lambda)$ and $p(a|x, y, \lambda) = p(a|x, \lambda)$ (similarly to b). Thus, any observable data – given by the probability distribution $p(a, b|x, y)$ – compatible with this LHV model can be decomposed as

$$p(a, b|x, y) = \sum_{\lambda} p(a|x, \lambda)p(b|y, \lambda)p(\lambda). \quad (1)$$

That is, any local distribution must lie inside the convex set defined by (1), the so-called correlation polytope \mathbb{C} [35, 36]. In this geometric picture, (linear) Bell inequalities are nothing else than facets of \mathbb{C} . Given that it is easy to list the extremal points of \mathbb{C} , to find its facets amounts to an efficient linear program, arguably the

reason why this method has become the most prominent in the study of nonlocality.

Another equivalent, but far less used method, comes from the realization that Bell inequalities are constraints arising from a marginal problem [37–39], that can be stated as: given some marginal distributions of n variables is it possible to find a joint distribution of all variables, such that this distribution marginalizes to the given ones? To see that Bell's theorem is indeed a particular marginal problem, notice that the LHV description (1) is equivalent to the existence of a joint distribution $\mathbf{p} = p(a_0, \dots, a_{m_x}, b_0, \dots, b_{m_y})$ (represented as a vector \mathbf{p}) describing the probability for outcomes of all possible measurements, where a_i labels the outcome a given that $x = i = \{0, \dots, m_x\}$ and similarly for b . Since \mathbf{p} defines a valid probability, it is constrained by a set of linear inequalities $L\mathbf{p} \geq 0$ given by $p_i \geq 0$ (positivity) and $\sum_i p_i = 1$ (normalization) defining the simplex polytope \mathbb{P} [40]. Given that at each round of the experiment only one a_i and one b_j can be measured simultaneously, \mathbf{p} defines a non-observable quantity. However, the constraints on \mathbf{p} will also imply constraints on the level of the observable distributions $p(a_i, b_j)$. These are exactly Bell inequalities, that in this picture can be understood as a condition for the marginal problem to have a positive answer. Thus, to obtain Bell inequalities in this picture, we have to eliminate from our description all non-observable terms. This is achieved via a FM elimination [33], a standard algorithm for the elimination of variables from a system of inequalities.

For simplicity and without loss of generality, in the remaining of the paper we focus on dichotomic outcomes (e.g. $a_i = 0, 1$). It is then convenient to consider the equivalent description of the problem in terms of the correlation vector \mathbf{E} with components given by expectation values, e.g., $\langle A_i B_j \rangle = \sum_{a_i, b_j} (-1)^{a_i + b_j} p(a_i, b_j)$. The vectors \mathbf{E} and \mathbf{p} are linearly related as $\mathbf{E} = T^{-1}\mathbf{p}$ implying that \mathbf{E} must obey linear inequalities $T\mathbf{E} \geq 0$ plus a normalization constraint. To illustrate the FM elimination consider the CHSH scenario [41] where each of the two parties in Fig. 1(a) can measure two observables. The inequalities

$$\langle A_0 B_0 \rangle + \langle A_1 B_0 \rangle - \langle A_0 A_1 \rangle \leq 1, \quad (2)$$

$$\langle A_0 B_1 \rangle - \langle A_1 B_1 \rangle + \langle A_0 A_1 \rangle \leq 1, \quad (3)$$

directly follow from $T\mathbf{E} \geq 0$ after the elimination of terms like $\langle A_0 A_1 B_0 B_1 \rangle$ and $\langle A_0 A_1 B_0 \rangle$. The sum of (2) and (3) eliminates the remaining non-observable term $\langle A_0 A_1 \rangle$ leading exactly to the CHSH inequality [41].

II. POLYNOMIAL BELL INEQUALITIES

Similarly to LHV models, a GLHV model also implies the existence of a well defined a joint distribu-

tion \mathbf{p} characterized by linear inequalities $L\mathbf{p} \geq 0$. The difference resides on the fact that GLHV models also imply a set of non-linear inequalities $W\mathbf{p} \geq 0$ (where $W = W(\mathbf{p})$). Thus, a GLHV model is characterized by intersection of \mathbb{P} with $W\mathbf{p} \geq 0$, that is, a semi-algebraic set [24]. As discussed before, this system of inequalities involves non-observable quantities that have to be eliminated in order to obtain a description in terms of empirically accessible variables only. Formally, the problem at hand is equivalent to a quantifier elimination: the projection of a semi-algebraic set onto a subspace of it, that by Tarski-Seidenberg theorem is again guaranteed to be a semi-algebraic set [24]. In other terms, the correlations compatible with a GLHV model are characterized by finitely many polynomial Bell inequalities. Quantifier elimination is routinely encountered in algebraic geometry problems, thus general purpose methods have been developed [24]. Unfortunately, given their computational complexity, their application to Bell scenarios is intractable even for the simplest possible models [4]. Notwithstanding, we show next that a simple adaptation of the FM elimination leads to practical and computational tractable way for deriving polynomial Bell inequalities.

The class of DAGs we consider are those which display (conditional) independencies on the level of the joint distribution \mathbf{p} . This is the case, for instance, in the DAG of Fig. 1(b) implying the independence relation $p(a,c) = p(a)p(c)$ and for many other relevant scenarios in quantum information [5, 20, 21, 23, 29, 31, 32]. The method to derive polynomial inequalities for this class of scenarios proceed as follows.

Given \mathbf{p} we first need to list all its components that are to be eliminated from our description: \mathbf{p}_O and \mathbf{p}_{NO} stand, respectively, to the set of observable and non-observable (to be eliminated) components p_i . We also list all the terms in \mathbf{p}_{NO} appearing in a non-linear fashion in $W\mathbf{p} \geq 0$, labeled by \mathbf{p}_{WNO} . Notice that all terms appearing in \mathbf{p}_{NO} but not in \mathbf{p}_{WNO} can be eliminated via a usual FM elimination over $L\mathbf{p} \geq 0$, obtaining a new set of linear relations $L'\mathbf{p} \geq 0$. The terms in \mathbf{p}_{WNO} have to be eliminated considering $L'\mathbf{p} \geq 0$ and $W\mathbf{p} \geq 0$ jointly. To that aim, notice that $W\mathbf{p} \geq 0$ can be linearized by considering some of the variables as free parameters of the problem. Given $W\mathbf{p} \geq 0$ there is going to be a minimum set of variables \mathbf{p}'_{WNO} that need to be set to free parameters in order to linearize the problem. This means that we can apply a FM elimination to the remaining terms obtaining a final set of inequalities that will depend linearly on the observable terms \mathbf{p}_O and polynomially on terms \mathbf{p}'_{WNO} . The observable data will also imply linear constraints on the parameters \mathbf{p}'_{WNO} . Together with these constraints, the obtained polynomials can be further simplified by usual quantifier elimination methods, finally arriving at polynomial inequal-

ities involving observable data only. We highlight that following this procedure, one can derive all polynomial Bell inequalities following from the intersection of $L\mathbf{p} \geq 0$ and $W\mathbf{p} \geq 0$, that is, our method provides a full characterization of the GLHV models under consideration. In practice, however, a partial characterization (e.g. in terms of full correlators only) will often be the only computationally tractable approach.

To illustrate the general method, we start considering the bilocality scenario, one of few cases for which polynomial Bell inequalities are known [29–31]. The scenario involves three parties with correlations mediated via two independent sources (see Fig. 1(b)). In the particular case of two dichotomic measurements per party, the following inequality has been proven to hold [29–31]

$$\sqrt{|I|} + \sqrt{|J|} \leq 2, \quad (4)$$

where $I = \sum_{x,z=0,1} \langle A_x B_0 C_z \rangle$ and $J = \sum_{x,z=0,1} (-1)^{x+z} \langle A_x B_1 C_z \rangle$. However, the methods in [30, 31] cannot be easily generalized to different scenarios, for instance, considering three measurements per party. Next we show in details how our framework can be employed to easily prove (4). We then proceed to derive new polynomial Bell inequalities.

To derive (4), we consider the independence constraint following from the DAG in Fig. 1(b):

$$\langle A_0 A_1 C_0 C_1 \rangle = \langle A_0 A_1 \rangle \langle C_0 C_1 \rangle. \quad (5)$$

We need to combine (5) via a FM elimination with the linear inequalities $T\mathbf{E} \geq 0$. It is sufficient to consider two inequalities following from $T\mathbf{E} \geq 0$:

$$\pm I - \langle A_0 A_1 \rangle - \langle C_0 C_1 \rangle - \langle A_0 A_1 C_0 C_1 \rangle \leq 1, \quad (6)$$

$$\pm J + \langle A_0 A_1 \rangle + \langle C_0 C_1 \rangle - \langle A_0 A_1 C_0 C_1 \rangle \leq 1. \quad (7)$$

Substituting (5) in (6) and (7) and after some algebraic manipulations, we can combine both inequalities into a single polynomial inequality

$$2 \langle A_0 A_1 \rangle^2 + (\pm I \mp J) \langle A_0 A_1 \rangle - (\pm I \pm J + 2) \leq 0. \quad (8)$$

As discussed before, we arrive at an inequality that depends linearly on the observable data (terms I and J) but have a non-linear dependence on non-observable terms, in this case $\langle A_0 A_1 \rangle$. The minimum of the polynomial in (8) is achieved at $\langle A_0 A_1 \rangle = (\pm I \mp J)/4$, implying the inequality in terms of observable data only

$$-(1/8)(\pm I - \mp J)^2 - (\pm I \pm J + 2) \leq 0. \quad (9)$$

This is a quadratic equation that can be easily solved, e.g. for I , leading exactly to (4).

Another nice feature of our construction is the fact that independencies are not required to hold exactly.

For instance, we may be interested in quantifying how much a given constraint must be relaxed in order to explain some experimental data [11, 42]. In the bilocality scenario if we allow for correlations $\mathcal{C}_{AC} \geq |\langle A_0 A_1 C_0 C_1 \rangle - \langle A_0 A_1 \rangle \langle C_0 C_1 \rangle|$ between parts A and C, it follows that

$$-(1/8)(\pm I - \mp J)^2 - (\pm I \pm J + 2) \leq 2\mathcal{C}_{AC}, \quad (10)$$

that is, the violation of (9) quantifies the degree of correlation required to classically reproduce some non-bilocal correlation. As an illustration, consider the correlation $I = J = 2$ that can be achieved quantum mechanically with two copies of Bell states shared between the parties [31]. In order to be classically reproduced, this correlation requires $\mathcal{C}_{AC} = 1$, that is, maximal correlation between parts A and C.

To further illustrate the practicality and relevance of our method we also derived new polynomial inequalities. See the Appendix for a detailed discussion. For the considerably more complicated GLHV model in Fig. 1(c) the inequality (9) is also valid if we define new functions given by $I = -\langle A_1 B_0 C_0 D_0 \rangle - \langle A_1 B_0 C_0 D_1 \rangle + \langle A_1 B_1 C_0 D_0 \rangle + \langle A_1 B_1 C_0 D_1 \rangle$ and $J = \langle A_0 B_0 C_1 D_0 \rangle - \langle A_0 B_0 C_1 D_1 \rangle + \langle A_0 B_1 C_1 D_0 \rangle - \langle A_0 B_1 C_1 D_1 \rangle$. Considering the bilocality scenario in Fig. 1(b) with 3 measurement settings, the following inequality holds:

$$-(1/8)(I - J + 16)^2 + 8I \leq 0, \quad (11)$$

with $I = \sum_{x,z=0,1,2} \langle A_x B_0 C_z \rangle$ and $J = \sum_{x,z=0,1,2} (-1)^{x+z} \langle A_x B_1 C_z \rangle$. To show the relevance of this inequality, notice that without the independence constraint it follows that $|I| + |J| \leq 10$. Choosing a correlation given $I = J = 9v$ (achievable in quantum mechanics for $v \leq 1/2$) we see that only for $v > 5/9$ the correlation is nonlocal. However, using (11) we see that this correlation is non-bilocal for $v > 4/9$, illustrating the gap between the local and bilocal sets.

III. NONSIGNALLING CORRELATIONS AND GENERALIZED BAYESIAN NETWORKS

In the study of nonlocality it is often useful to define the notion of nonsignalling (NS) correlations [43]. These are the observable distributions \mathbf{p}_O that cannot be used to signal between the parties, that is, the marginal distributions are well defined quantities that cannot depend in any way on which observable the other parties have measured. A paradigmatic example of a NS correlation is the Popescu-Rohrlich(PR)-box defined as $p(a, b|x, y) = 1/2\delta_{a\oplus b, xy}$. [34].

The marginal problem approach naturally incorporates the notion of NS correlations [34]. For instance, for the bipartite scenario in Fig. 1(a), NS correlations

are those that have well defined observable distributions $p(a_i, b_j) \forall i, j$ (respecting positivity and normalization) and well defined marginals, that is, $p(a_i) = \sum_{b_j} p(a_i, b_j) = \sum_{b_{j'}} p(a_i, b_{j'}) \forall j, j'$ (and similarly for $p(b_j)$). These constraints can be combined into a system of linear inequalities $L_{NSPO} \geq 0$ defining a polytope that is characterized by finitely many extremal points.

We can define NS correlations in generalized Bayesian networks, as those that are compatible with GLHV models where all the underlying (classical) hidden variables are replaced by general NS distributions. As shown in [9], all the conditional independencies on the level of the observable distributions \mathbf{p}_O that are valid in the classical setup will remain valid after this replacement. For instance, for the bilocality scenario in Fig. 1(b), even if we allow for NS correlations to be shared between the parties, it is true that the statistics between parties A and C should factorize, that is, $p(a_i, c_k) = p(a_i)p(c_k)$. As before, we can represent the observable independencies as a system of polynomial inequalities $W_{NSPO} \geq 0$. Thus, we can define generalized nonsignalling (GNS) correlations as those inside the semi-algebraic set Σ , defined by intersection of $L_{NSPO} \geq 0$ and $W_{NSPO} \geq 0$. Since Σ defines a non-convex body characterized by polynomial inequalities, differently from the usual case, there are going to be infinitely many extremal GNS points defining it. In spite of that, we can still define a sensible and practical way to characterize GNS correlations. Similarly to what has been done before, we can take some of the variables appearing in $W_{NSPO} \geq 0$ as free parameters in order to linearize it. Doing that we turn Σ into a convex set with finitely many extremal points that can therefore be characterized by standard linear program techniques.

As an illustration consider the GLHV model in Fig. 1(b) with all parties performing two dichotomic measurements. If we fix the marginal distributions of parts A and C to be $p(a_i) = p(c_k) = 1/2$, we see that one of the extremal GNS points is given by $p(a, b, c|x, y, z) = 1/4\delta_{a\oplus b\oplus c, y(x\oplus z)}$, a distribution that can be achieved replacing the hidden variables in Fig. 1(b) by PR-boxes [43].

IV. DISCUSSION

Bayesian networks offer an almost unexplored ground for generalizations of Bell's theorem. The basic question to be solved in this quest is how to derive polynomial Bell inequalities associated with more complex causal structures. In this work we made an important step in that direction. We proposed a practical and general method that can be readily applied to a wide range of scenarios, considering its applications in few GLHV models and deriving polynomial Bell inequalities char-

acterizing them. We have also shown how our construction naturally leads to a notion of nonsignalling correlations in GLHV models.

Given the fundamental role that Bell inequalities play in the study and practical applications of nonlocality, we believe that our results will motivate and set a basic tool for future research in generalized Bell scenarios. The natural next step is to put the machinery to use in a variety of scenarios and derive new Bell inequalities well suited, for example, to decrease the requirements on experimental implementations of Bell tests [44]. It would be interesting to investigate the role of polynomial Bell inequalities in practical applications of nonlocality, such as quantum cryptography [45], randomness generation [46, 47] or distributed computing [48]. For instance, the amount of violation of usual Bell inequalities can be directly associated with the probability of success in communication complexity problems [48, 49]. Are there any communication problems associ-

ated to polynomial Bell inequalities? Another possibility is to find Tsirelson's bounds [50, 51] associated with these generalized inequalities, that is, what is the maximum violation of them achievable with quantum correlations. Related to that and inspired by results such as information causality [52], it would also be relevant to derive information-theoretical principles for these more complex Bayesian networks [11].

ACKNOWLEDGMENTS

We acknowledge financial support from the Excellence Initiative of the German Federal and State Governments (Grants ZUK 43 & 81), the US Army Research Office under contracts W911NF-14-1-0098 and W911NF-14-1-0133 (Quantum Characterization, Verification, and Validation), the DFG (GRO 4334 & SPP 1798).

[1] J. S. Bell, *Physics* **1**, 195 (1964).

[2] J. Pearl, *Causality* (Cambridge University Press, 2009).

[3] P. Spirtes, N. Glymour, and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. (The MIT Press, 2001).

[4] G. Ver Steeg and A. Galstyan, in *Proceedings of the 27th conference on Uncertainty in Artificial Intelligence* (2011).

[5] T. Fritz, *New Journal of Physics* **14**, 103001 (2012).

[6] T. Fritz, arXiv preprint arXiv:1404.4812 (2014).

[7] R. Chaves, L. Luft, T. O. Maciel, D. Gross, D. Janzing, and B. Schölkopf, *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, 112 (2014).

[8] C. J. Wood and R. W. Spekkens, *New Journal of Physics* **17**, 033002 (2015).

[9] J. Henson, R. Lal, and M. F. Pusey, *New Journal of Physics* **16**, 113043 (2014).

[10] R. Chaves, C. Majenz, and D. Gross, *Nature communications* **6**, 5766 (2015).

[11] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, *Phys. Rev. Lett.* **114**, 140403 (2015).

[12] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, *Nature Physics* **11**, 414 (2015).

[13] R. F. Werner and M. M. Wolf, *Phys. Rev. A* **64**, 032112 (2001).

[14] M. Zukowski and C. Brukner, *Phys. Rev. Lett.* **88**, 210401 (2002).

[15] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, *Phys. Rev. Lett.* **88**, 040404 (2002).

[16] D. Collins and N. Gisin, *Journal of Physics A: Mathematical and General* **37**, 1775 (2004).

[17] S. Popescu, *Phys. Rev. Lett.* **74**, 2619 (1995).

[18] R. Gallego, L. E. Würflinger, R. Chaves, A. Acín, and M. Navascués, *New Journal of Physics* **16**, 033037 (2014).

[19] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Rev. Mod. Phys.* **86**, 419 (2014).

[20] A. Acín, J. I. Cirac, and M. Lewenstein, *Nature Physics* **3**, 256 (2007).

[21] M. Zukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, *Phys. Rev. Lett.* **71**, 4287 (1993).

[22] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin, *Rev. Mod. Phys.* **83**, 33 (2011).

[23] A. Sen(De), U. Sen, C. Brukner, V. Buzek, and M. Zukowski, *Phys. Rev. A* **72**, 042310 (2005).

[24] D. Geiger and C. Meek, in *Proceedings of the 15th conference on Uncertainty in Artificial Intelligence* (1999) pp. 226–235.

[25] L. D. Garcia, M. Stillman, and B. Sturmfels, *Journal of Symbolic Computation* **39**, 331 (2005).

[26] B. Steudel and N. Ay, *Entropy* **17**, 2304 (2015).

[27] R. Chaves and T. Fritz, *Phys. Rev. A* **85**, 032113 (2012).

[28] R. Chaves, L. Luft, and D. Gross, *New J. Phys.* **16**, 043001 (2014).

[29] C. Branciard, N. Gisin, and S. Pironio, *Phys. Rev. Lett.* **104**, 170401 (2010), 1112.4502.

[30] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, *Phys. Rev. A* **85**, 032119 (2012).

[31] A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, *Phys. Rev. A* **90**, 062109 (2014).

[32] K. Mukherjee, B. Paul, and D. Sarkar, arXiv preprint arXiv:1411.4188 (2014).

[33] H. P. Williams, *Amer. Math. Monthly* **93**, 681 (1986).

[34] S. Popescu and D. Rohrlich, *Foundations of Physics* **24**, 379 (1994).

[35] I. Pitowsky, *Quantum probability–quantum logic*, Lecture notes in physics (Springer-Verlag, 1989).

[36] I. Pitowsky, *Mathematical Programming* **50**, 395 (1991).

[37] C. Budroni and A. Cabello, *Journal of Physics A: Mathematical and Theoretical* **45**, 385304 (2012).

[38] T. Fritz and R. Chaves, *IEEE Trans. Inform. Theory* **59**, 803 (2013).

[39] P. Kurzyński and D. Kaszlikowski, *Phys. Rev. A* **89**, 012103 (2014).

[40] S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge university press, 2009).

[41] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[42] M. J. W. Hall, Phys. Rev. Lett. **105**, 250404 (2010).

[43] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Phys. Rev. A **71**, 022101 (2005).

[44] P. H. Eberhard, Phys. Rev. A **47**, R747 (1993).

[45] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[46] R. Colbeck, *Ph.D. Thesis*, Ph.D. thesis, University of Cambridge (2007).

[47] S. Pironio *et al.*, Nature **464**, 1021 (2010).

[48] C. Brukner, M. Zukowski, J.-W. Pan, and A. Zeilinger, Phys. Rev. Lett. **92**, 127901 (2004).

[49] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Reviews of modern physics **82**, 665 (2010).

[50] B. Cirel'son, Letters in Mathematical Physics **4**, 93 (1980).

[51] M. Navascués, S. Pironio, and A. Acín, Phys. Rev. Lett. **98**, 010401 (2007).

[52] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, Nature **461**, 1101 (2009).

[53] W. Research, "Mathematica 8.0," (2010).

Appendix A: A method for the derivation of polynomial Bell inequalities

As discussed in the main text, the derivation of polynomial Bell inequalities follows from an adapted FM elimination over the combined system of inequalities $L\mathbf{p} \geq 0$ and $W\mathbf{p} \geq 0$ (with $W = W(\mathbf{p})$), the first representing linear relations respected by a well defined probability distribution \mathbf{p} while the latter stands for the (conditional) independence (CI) constraints implied by a given GLHV model. Notice, however, that not all DAGs will display CIs on the level of \mathbf{p} ; this is the case for instance in the so-called triangle scenario [5, 26, 28].

Given the scenario of interest, we need to define \mathbf{p}_O and \mathbf{p}_{NO} standing, respectively, to the set of components p_i that we want to keep or not in our description. We also need to define \mathbf{p}_{WNO} and \mathbf{p}'_{WNO} . The first corresponds to the components in \mathbf{p}_{NO} appearing in a non-linear fashion in the inequalities $W\mathbf{p} \geq 0$ while the latter describes the minimum set of components that need to be taken as free real parameters in order to linearize $W\mathbf{p} \geq 0$. All the terms in \mathbf{p}_{NO} but not in \mathbf{p}_{WNO} can be eliminated via a usual FM elimination leading to new set of inequalities $L'\mathbf{p} \geq 0$. To understand the FM elimination, notice that since the sum of two valid inequalities also defines a valid inequality, in order to eliminate a given term from our description we basically have to consider all possible pairwise sums of inequalities where the coefficients of the term to be eliminated appear with opposite signs. The remaining terms have to be eliminated resorting to the adapted FM method discussed in the main text and illustrated in details below.

As a side remark, we notice that instead of performing the usual FM elimination leading to $L'\mathbf{p} \geq 0$, one can equivalently list the extremal points in the subspace given by the support of L' , and then dualize the description in order to exactly obtain the inequalities $L'\mathbf{p} \geq 0$. In practice which approach will be better is going to depend on the scenario in question. Typically, if the number of terms in \mathbf{p}_O is not large, the dualization approach will be reasonably faster. We refer the reader to Ref. [37] for a discussion of the computational advantage of the both methods in usual LHV models. In the following, for simplicity and without loss of generality, we focus on the case where all measurements have dichotomic outcomes so that we can equivalently treat the problem in terms of the correlation vector \mathbf{E} with components given by expectation values.

To illustrate the abstract discussion, consider the bilocality scenario (see Fig. 1(b) in the main text) in the particular case where all the parties measure two dichotomic observables, implying the bilocality constraint $p_{ac}(a_0, a_1, c_0, c_1) = p_a(a_0, a_1)p_c(c_0, c_1)$. Since the variables are binary, we see that the bilocality assumption is equivalent to 16 (not necessarily independent) quadratic constraints. In order to linearize this set of constraints we can take $v_{a_0, a_1} = p_a(a_0, a_1)$ (for each $a_0, a_1 = 0, 1$) as a free real parameter, that is, we can express the non-linear constraints $W(\mathbf{p})\mathbf{p} \geq 0$ as a linear relation $W(v_{0,0}, v_{0,1}, v_{1,0}, v_{1,1})\mathbf{p} \geq 0$. In terms of expectation values, we have a correlation vector $\mathbf{E} = (1, E_{C_0}, E_{C_1}, E_{C_0C_1}, \dots, E_{A_0A_1B_0B_1C_0C_1})$ (for simplicity we label $\langle X \rangle = E_X$) with 64 components that must respect the linear constraints $T\mathbf{E} \geq 0$ (with $\mathbf{E} = T^{-1}\mathbf{p}$). The bilocality constraints can also be expressed in terms of expectation values. For instance, $p_{ac}(0, 0, 0, 0) = p_a(0, 0)p_c(0, 0)$ is equivalent to

$$v(1 + E_{C_1} + E_{C_0} + E_{C_0C_1}) + E_{A_0A_1} + E_{A_0} + E_{A_1} + E_{A_0C_0} + E_{A_0C_1} + E_{A_1C_0} + E_{A_1C_1} + E_{A_1C_0C_1} + E_{A_0C_0C_1} + E_{A_0A_1C_1} + E_{A_0A_1C_0} + E_{A_0A_1C_0C_1} = 0 \quad (A1)$$

with $v = 1 - 4p_a(0, 0)$, where $p_a(0, 0)$ is the real free parameter. Notice that the bilocality constraints do not depend on the variables B_0 and B_1 . Therefore all non-observable terms that depend on them, for instance, $E_{A_0A_1B_0B_1C_0C_1}$, can be eliminated via the usual FM elimination method over $T\mathbf{E} \geq 0$, defining a new system of linear inequalities $T'\mathbf{E} \geq 0$. The remaining terms to be eliminated are those that depend jointly on A_0, A_1 and/or C_0, C_1 . We further notice that all the bilocal constraints (e.g. (A1)) only have a non-linear dependence on the terms E_{C_0} ,

E_{C_1} and $E_{C_0C_1}$. That is, all the non-observable terms but $E_{C_0C_1}$ can be eliminated via the usual FM elimination. After all non-observable terms have been eliminated we arrive at a final description that depends linearly on the observable terms and non-linearly on the free parameters v_{a_0, a_1} . Notice that the observable data will also imply linear constraints on the free parameters themselves. Together with these constraints, each of obtained polynomial inequalities can be further simplified by usual quantifier elimination methods (see for instance the function *Reduce* in Mathematica [53]), finally arising at polynomial inequalities involving the observable data only.

In the following we will apply the general method to each of the scenarios in Fig. 1 of the main text. For computational reasons we consider the case with full correlators but no marginal terms, that is, we keep terms like $\langle A_i B_j C_k \rangle$ but not terms like $\langle A_i B_j \rangle$ or $\langle A_i \rangle$. Notwithstanding, as highlighted in the main text, our method can also be applied to obtain a full characterization of the GLHV models, that is, including marginal terms.

Appendix B: Detailed derivation of the polynomial Bell inequalities

We start considering the bilocality scenario in Fig. 1(b) in the case where each party A, B and C can measure two dichotomic observables. We highlight that the same analysis remains valid if we consider part B to perform a single measurement with 4 possible outcomes, e.g., a measurement in the Bell basis. Restricting to the subspace of full correlators –that is, containing terms $\langle A_i B_j C_k \rangle$ – we observe that one of the obtained inequalities is exactly eq. (8) of the main text. That is, this inequality corresponds to a facet of the bilocal set in the subspace of full correlators. We further notice that in order to derive this class of inequalities it is sufficient to consider – instead of the full set of bilocal constraints – the simple constraint

$$\langle A_0 A_1 C_0 C_1 \rangle = \langle A_0 A_1 \rangle \langle C_0 C_1 \rangle. \quad (\text{B1})$$

Together with the linear constraints

$$\pm I - \langle A_0 A_1 \rangle - \langle C_0 C_1 \rangle - \langle A_0 A_1 C_0 C_1 \rangle \leq 1, \quad (\text{B2})$$

$$\pm J + \langle A_0 A_1 \rangle + \langle C_0 C_1 \rangle - \langle A_0 A_1 C_0 C_1 \rangle \leq 1, \quad (\text{B3})$$

we can readily prove inequality eq. (8) of the main text. Inequalities (B2) and (B3) directly follow from $TE \geq 0$ with $I = \sum_{x,z} \langle A_x B_0 C_z \rangle$ and $J = \sum_{x,z} (-1)^{x+z} \langle A_x B_1 C_z \rangle$. Substituting (B1) in (B2) and (B3) we obtain

$$\pm I - (1 + \langle A_0 A_1 \rangle) - \langle C_0 C_1 \rangle (1 + \langle A_0 A_1 \rangle) \leq 0, \quad (\text{B4})$$

$$\pm J - (1 - \langle A_0 A_1 \rangle) + \langle C_0 C_1 \rangle (1 - \langle A_0 A_1 \rangle) \leq 0. \quad (\text{B5})$$

Notice that $(1 \pm \langle A_0 A_1 \rangle) \geq 0$, with equality only if $\langle A_0 A_1 \rangle = \mp 1$, that is, only if both outputs of part A are deterministic functions, that is, only if we have the trivial case where either $I = 0$ or $J = 0$. For $(1 \pm \langle A_0 A_1 \rangle) > 0$, we can rearrange (B4) and (B5) as

$$\pm I / (1 + \langle A_0 A_1 \rangle) - 1 - \langle C_0 C_1 \rangle \leq 0, \quad (\text{B6})$$

$$\pm J / (1 - \langle A_0 A_1 \rangle) - 1 + \langle C_0 C_1 \rangle \leq 0. \quad (\text{B7})$$

Summing both inequalities we eliminate the term $\langle C_0 C_1 \rangle$ and arrive at

$$\pm I / (1 + \langle A_0 A_1 \rangle) \pm J / (1 - \langle A_0 A_1 \rangle) \leq 2, \quad (\text{B8})$$

that can be further arranged to obtain the class of inequalities discussed in the main text, given by

$$2 \langle A_0 A_1 \rangle^2 + (\pm J \mp I) \langle A_0 A_1 \rangle - (2 \pm I \pm J) \leq 0. \quad (\text{B9})$$

As discussed before, we arrive at an inequality that depends linearly on the observable data (terms I and J) but have a non-linear dependence on the non-observable term $\langle A_0 A_1 \rangle$. Given I and J , to check that this data fulfills the inequality, we have to prove that there is at least one choice of $\langle A_0 A_1 \rangle$ such that the lhs of (B9) is ≤ 0 . That is, to make use of (B9) we have to find the value of $\langle A_0 A_1 \rangle$ (as a function of I and J) minimizing the polynomial on the lhs. Since the lhs in (B9) defines a convex function, the minimum of the inequality is achieved at $\langle A_0 A_1 \rangle = (\pm I \mp J) / 4$, implying the inequality in terms of observable data only.

$$-(1/8)(\pm I \mp J)^2 - (2 \pm I \pm J) \leq 0. \quad (\text{B10})$$

A similar derivation is possible if we allow for correlation between parts A and C, such that $|\langle A_0 A_1 C_0 C_1 \rangle - \langle A_0 A_1 \rangle \langle C_0 C_1 \rangle| \leq \mathcal{C}_{AC}$. Summing $\langle A_0 A_1 C_0 C_1 \rangle - \langle A_0 A_1 \rangle \langle C_0 C_1 \rangle \leq \mathcal{C}_{AC}$ with (B2) and (B3) we obtain

$$\pm I - (1 + \langle A_0 A_1 \rangle) - \langle C_0 C_1 \rangle (1 + \langle A_0 A_1 \rangle) \leq \mathcal{C}_{AC}, \quad (\text{B11})$$

$$\pm J - (1 - \langle A_0 A_1 \rangle) + \langle C_0 C_1 \rangle (1 - \langle A_0 A_1 \rangle) \leq \mathcal{C}_{AC}. \quad (\text{B12})$$

Proceeding with the exact same steps as above we finally obtain

$$-(1/8)(\pm I \mp J)^2 - (2 \pm I \pm J) \leq 2\mathcal{C}_{AC}. \quad (\text{B13})$$

To prove that a similar inequality holds for the GLHV model in Fig. 1(c) we can follow a very similar derivation. For this model it follows the independence constraint

$$\langle B_0 B_1 D_0 D_1 \rangle = \langle B_0 B_1 \rangle \langle D_0 D_1 \rangle. \quad (\text{B14})$$

Together with the linear constraints

$$\pm I - \langle B_0 B_1 \rangle + \langle D_0 D_1 \rangle + \langle B_0 B_1 D_0 D_1 \rangle \leq 1, \quad (\text{B15})$$

$$\pm J + \langle B_0 B_1 \rangle - \langle D_0 D_1 \rangle + \langle B_0 B_1 D_0 D_1 \rangle \leq 1, \quad (\text{B16})$$

where $I = -\langle A_1 B_0 C_0 D_0 \rangle - \langle A_1 B_0 C_0 D_1 \rangle + \langle A_1 B_1 C_0 D_0 \rangle + \langle A_1 B_1 C_0 D_1 \rangle$ and $J = +\langle A_0 B_0 C_1 D_0 \rangle - \langle A_0 B_0 C_1 D_1 \rangle + \langle A_0 B_1 C_1 D_0 \rangle - \langle A_0 B_1 C_1 D_1 \rangle$, we can follow the same steps as above to prove that (B10) also holds in this scenario.

We now move to the scenario in Fig. 1(b) where parties A and C can measure three possible observables. To prove that inequality (11) of the main text holds in this case, we need to proceed as follows. We have to consider the inequalities

$$4J - 3\langle f_1 \rangle + \langle f_1 C_0 C_1 \rangle - \langle f_1 C_0 C_2 \rangle + \langle f_1 C_1 C_2 \rangle \leq 0, \quad (\text{B17})$$

$$4I - 3\langle f_2 \rangle - \langle f_2 C_0 C_1 \rangle - \langle f_2 C_0 C_2 \rangle - \langle f_2 C_1 C_2 \rangle \leq 0, \quad (\text{B18})$$

that follow from $TE \geq 0$ with $f_1 = +3 - A_0 A_1 + A_0 A_2 - A_1 A_2$, $f_2 = 3 + A_0 A_1 + A_0 A_2 + A_1 A_2$. It also follows from $TE \geq 0$ that

$$3f_2 \geq 2|I|, \quad (\text{B19})$$

$$3f_1 \geq 2|J|, \quad (\text{B20})$$

$$f_1 + f_2 \leq 8. \quad (\text{B21})$$

Using the independence relation

$$\langle A_i A_j C_k C_l \rangle = \langle A_i A_j \rangle \langle C_k C_l \rangle \quad \forall i, j, k, l \quad (\text{B22})$$

we can rewrite (B17) and (B18) as

$$4J + \langle f_1 \rangle (-3 + \langle C_0 C_1 \rangle - \langle C_0 C_2 \rangle + \langle C_1 C_2 \rangle) \leq 0, \quad (\text{B23})$$

$$4I + \langle f_2 \rangle (-3 - \langle C_0 C_1 \rangle - \langle C_0 C_2 \rangle - \langle C_1 C_2 \rangle) \leq 0. \quad (\text{B24})$$

Since (B19) and (B20) imply that f_1 and f_2 are strictly positive quantities (apart from the trivial case $I = 0$ and/or $J = 0$), we can rewrite these inequalities as

$$4J/f_1 + (-3 + \langle C_0 C_1 \rangle - \langle C_0 C_2 \rangle + \langle C_1 C_2 \rangle) \leq 0, \quad (\text{B25})$$

$$4I/f_2 + (-3 - \langle C_0 C_1 \rangle - \langle C_0 C_2 \rangle - \langle C_1 C_2 \rangle) \leq 0. \quad (\text{B26})$$

$$(\text{B27})$$

Summing both inequalities we eliminate the terms $\langle C_0 C_1 \rangle$ and $\langle C_1 C_2 \rangle$, obtaining

$$4I/f_2 + 4J/f_1 - 2\langle C_0 C_2 \rangle \leq 6. \quad (\text{B28})$$

Combining it with the trivial inequality $\langle C_0 C_2 \rangle \leq 1$, we finally obtain

$$I f_1 + J f_2 \leq 2f_1 f_2. \quad (\text{B29})$$

Since (B19) implies that $\pm I - 2f_2$ is a strictly negative quantity (apart from the trivial case $|I| = 2f_2$), we can rewrite the inequality above as

$$-f_1 - Jf_2/(I - 2f_2) \leq 0. \quad (\text{B30})$$

Summing it with (B19), we obtain

$$f_2 - Jf_2/(I - 2f_2) \leq 8, \quad (\text{B31})$$

that can be rewritten as

$$2f_2^2 + f_2(-I + J - 16) + 8I \leq 0. \quad (\text{B32})$$

The lhs is a quadratic equation on the non-observable term that is convex, implying that the minimum of the lhs is obtained at $f_2 = (+I - J + 16)/4$ and therefore

$$-(1/8)(+I - J + 16)^2 + 8I \leq 0. \quad (\text{B33})$$

